Title: Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system

Authors: Tiffany B. Taylor[1†], Geraldine Mulley[1†], Alexander H. Dills[2], Abdullah S. Alsohim[1,3], Liam J. McGuffin[1], David J. Studholme[4], Mark W. Silby[2], Michael A. Brockhurst[5], Louise J. Johnson[1,*], Robert W. Jackson[1,6]

Affiliations:

[1]School of Biological Sciences, University of Reading, Whiteknights, Reading, RG6 6AJ, UK

[2]Department of Biology, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, MA 02747, USA

[3]Department of Plant Production and Protection, Qassim University, Qassim, Saudi Arabia P.O. Box 6622

[4]College of Life and Environmental Sciences, University of Exeter, Stocker Road, Exeter, EX4 4QD, UK

[5]Department of Biology, University of York, Wentworth Way, York, YO10 5DD, UK

[6]The University of Akureyri, Borgir vid Nordurslod, IS-600 Akureyri, Iceland

*Corresponding author: L.J.Johnson@reading.ac.uk

† These authors contributed equally to this work

One Sentence Summary:

Rapid repeatable rewiring of regulatory networks: a nitrogen regulatory gene evolves a new function, restoring flagella to immotile bacteria.

Abstract:

A central process in evolution is the recruitment of genes to regulatory networks. We engineered immotile strains of the bacterium *Pseudomonas fluorescens* that lack flagella due to deletion of the regulatory gene *fleQ*. Under strong selection for motility, these bacteria consistently regained flagella within 96 hours via a two-step evolutionary pathway. Step 1 mutations increase intracellular levels of phosphorylated NtrC, a distant homologue of FleQ*,* which begins to commandeer control of the *fleQ* regulon at the cost of disrupting nitrogen uptake and assimilation. Step 2 is a switch-of-function mutation that redirects NtrC away from nitrogen uptake and towards its novel function as a flagellar regulator. Our results demonstrate that natural selection can rapidly rewire regulatory networks in very few, repeatable mutational steps.

Main Text:

A longstanding evolutionary question concerns how the duplication and recruitment of genes to regulatory networks facilitates their expansion (*1*), and how networks gain mutational robustness and evolvability (*2*). Bacteria respond to diverse environments using a vast range of specialised regulatory pathways, predominantly two-component systems (*3*), which are the result of adaptive radiations within gene

families. Due to past cycles of gene duplication, divergence and horizontal genetic transfer, there is often extensive homology between the components of different pathways (*4*), raising the possibility of cross-talk or redundancy between pathways (*5*). Here we monitor the recovery of microbial populations from a catastrophic gene deletion: bacteria engineered to lack a particular function are exposed to environments that impose strong selection to evolve it *de novo* (*6, 7, 8, 9*). We use this approach to understand the recruitment of genes to regulatory networks in real-time.

In the plant-associated soil bacterium *P. fluorescens,* the master regulator of flagellar synthesis is FleQ (also called AdnA), a σ^{54}-dependent enhancer binding protein (EBP) that activates transcription of genes required for flagellum biosynthesis (*10, 11*). The starting *P. fluorescens* strain is AR2; this strain lacks flagella, due to deletion of *fleQ,* and is unable to move by spreading motility due to mutation of viscosin synthase (*viscB*), resulting in a distinctive, point-like colony morphology on spreading motility medium (SMM) (*12*) (Figure 1A). We grew replicate populations of AR2 on SMM; when local nutrients became depleted, starvation imposed strong selection to re-evolve motility. To demonstrate that this finding was not strain-specific, these experiments were replicated in a different strain of *P. fluorescens,* Pf0-2x. This strain is a Δ*fleQ* variant of Pf0-1, already viscosin-deficient, and is thus unable to move by spreading or swimming motility.

After 96 hours incubation of AR2 and Pf0-2x at room temperature on SMM, two breakout mutations were visible conferring first slow (AR2S and Pf0-2xS) and then fast (AR2F and Pf0-2xF) spreading over the agar surface (Fig. 1A). The AR2F strain produces flagella, but we could not detect flagella in EM samples for AR2S (Fig. 1B).

Genome resequencing revealed a single nucleotide point mutation in *ntrB* in strain AR2S, causing an amino acid substitution within the PAS domain of the histidine kinase (HK) sensor NtrB (T97P). The fast-spreading strain AR2F had acquired an additional point mutation in the σ^{54}-dependent EBP gene *ntrC,* which alters an amino acid (R442C) within the DNA-binding domain (Table 1 & S2).

NtrB and NtrC comprise a two-component system: under nitrogen limitation NtrB phosphorylates NtrC, which activates transcription of genes required for nitrogen uptake and metabolism. To determine how mutations in this separate regulatory pathway restored motility in the absence of FleQ, we performed microarray and qRT-PCR analyses of the ancestral and evolved strains (Fig. S1 & Table S1). The expression of genes required for flagellum biosynthesis and chemotaxis was abolished in AR2 compared to wild-type SBW25 (Fig. 2A). The *ntrB* mutation in AR2S partially restores the expression of flagellar genes, and over-activates the expression of genes involved in nitrogen regulation, uptake and metabolism. The subsequent *ntrC* mutation in AR2F reduces the expression of nitrogen uptake and metabolism genes, while further up-regulating flagellar and chemotaxis gene expression to wild-type levels (Fig. 2B). While AR2S and AR2F showed higher growth rates than the ancestor in LB medium (the medium on which the mutants arose; Tukey-Kramer HSD test, growth in LB compared to AR2: AR2S, $P < 0.001$; AR2F, $P < 0.001$) (Fig. 1C), both mutants grew poorly in minimal medium with ammonium as the sole nitrogen source (Tukey-Kramer HSD test, growth in M9 + ammonium compared to AR2: AR2S, $P < 0.001$; AR2F, $P = 0.001$). This is likely to be the result of ammonium toxicity due to the strong up-regulation of genes

involved in ammonium uptake and assimilation, indicating a pleiotropic cost of this adaptation.

Sequencing of the *ntrBC* locus from independently evolved replicate strains showed that evolution often followed parallel trajectories in both AR2 and Pf0-2x: mutation of *ntrB* gave a slow-spreading strain and this was followed by mutation of *ntrC* yielding a fast-spreading strain (Table 1; Fig. 3A & C). While all Pf0-2xF replicates carried mutations in *ntrC*, several Pf0-2xS strains were not mutated in *ntrB* suggesting an alternative evolutionary pathway. Genome resequencing of these strains revealed mutations in *glnA* or *glnK* (Table 1; Fig. 3B) likely to result in loss of function leading to abnormally high levels of phosphorylated NtrC: reduced ammonium assimilation by glutamine synthetase (*glnA*) would impose severe nitrogen limitation in the cell irrespective of nitrogen availability, whereas GlnK is a PII-protein that regulates both NtrB and glutamine synthetase activities.

These data suggest a predictable two-step evolutionary process: **Step 1** increases levels of phosphorylated NtrC, via either (a) a direct regulatory route with mutations in NtrB or GlnK, or (b) a physiological route with loss-of-function mutations reducing glutamine synthetase activity and causing NtrB activation, to partially or intermittently reactivate the flagellar cascade. In **Step 2**, NtrC adapts to enhance activation of the flagellar genes and in doing so, becomes a less potent activator of nitrogen uptake genes. This model explains the microarray data and is consistent with the predicted structural effects of the mutations (Figs. 2 and 3). Specifically, for NtrB the structural changes are likely to either increase kinase or reduce phosphatase activity. In

support of this, the mutated NtrB(D228A) repeatedly emerging in Pf0-2xS resembles NtrB(D227A) in *P. aeruginosa* which constitutively activates the Ntr system (*13*).

NtrC is a distant homologue of FleQ, sharing 30% amino acid identity and the same three structural domains (TM-score > 0.7; $P < 0.001$; Fig. 3D) (*14*): an N-terminal receiver domain, a conserved central σ^{54}-interacting domain, and a C-terminal DNA-binding domain containing a helix-turn-helix (HTH) motif flanked by highly disordered regions. We posit that an overabundance of phosphorylated NtrC activates transcription of flagellar genes through functional promiscuity (*15*). Consistent with this, the *ntrC* mutations in fast-spreading strains are predominantly located within or adjacent to the HTH domain and likely influence enhancer-binding specificity; one is a frameshift abolishing the HTH altogether (Fig. 3C). The evolved NtrC' must be constitutively phosphorylated by overactive NtrB to enable its new multipurpose role, with the result that flagellum biosynthesis and nitrogen regulation are probably no longer responsive to environmental stimuli. Consequently, there is a trade-off between nitrogen utilization and motility (Fig. 1A & C).

The flagellar regulatory network may have an unusually dynamic evolutionary history. Flagella are expensive to make, and not always advantageous. Pathogens expressing flagella can trigger strong immune responses in the host, so rapid transitions are seen over short timescales between uniflagellate, multiflagellate and aflagellate states (*16*). This volatility is reflected in the structure of regulatory networks: in close relatives of *Pseudomonas*, *fleQ* appears not to be involved in flagellar gene expression (*17*), and in *Helicobacter pylori*, a gene of unknown function can be used as a "spare part" to permit motility in *flhB* mutants (*18*). Our results illustrate that trans-acting

mutations can be important in gene network evolution (*19*), but that as predicted, such mutations bear severe pleiotropic costs (*20, 21*). Genes can retain the potential to take on the functions of long-diverged homologues, suggesting that some degree of evolutionary resilience is a consequence of regulatory pathways that evolve via gene duplication. The elegance of this system enables us to understand the adaptive process in detail at the genetic and phenotypic level, however it is worth noting that de novo origination of new functions in nature is likely to take longer and involve more mutational steps. Here we identified a novel, tractable model for gene network evolution and observed, in real time, the rewiring of gene networks to enable the incorporation of a modified component (NtrC') creating a novel regulatory function, by a highly repeatable two-step evolutionary pathway with the same point mutations often recurring in independent lineages.

References:

1. J. R. True, S. B. Carroll, Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell. Dev. Bi.* **18**, 53-80 (2002).

2. M. Aldana, E. Balleza, S. Kauffman, O. Resendiz, Robustness and evolvability in genetic regulatory networks. *J. Theor. Biol.* **245**, 433-448 (2007).

3. J. A. Hoch, Two-component and phosphorelay signal transduction. *Curr. Opin. Microbiol.* **3**, 165-170 (2000).

4. S. A. Teichmann, M. M. Babu, Gene regulatory network growth by duplication. *Nat. Genet.* **36**, 492-496 (2004).

5. E. J. Capra, M. T. Laub, Evolution of two-component signal transduction systems. *Annu. Rev. Microbiol.* **66**, 325-347 (2012).

6. Z. D. Blount, C. Z. Borland, R. E. Lenski, Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **105**, 7899-7906 (2008).

7. J. R. Meyer, *et al.*, Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335**, 428-432 (2012).

8. B. Hall, The EBG system of *E. coli*: Origin and evolution of a novel β-galactosidase for the metabolism of lactose. *Genetica* **118**, 143-156 (2003).

9. D. Blank, L. Wolf, M. Ackermann, O. K. Silander, The predictability of molecular evolution during functional innovation. *Proc. Natl. Acad. Sci. USA* **111**, 3044-3049 (2014).

10. E. A. Robleto, I. López-Hernández, M. W. Silby, S. B. Levy, Genetic analysis of the AdnA regulon in *Pseudomonas fluorescens*: Nonessential role of flagella in adhesion to sand and biofilm formation. *J. Bacteriol.* **185**, 453-460 (2003).

11. N. Dasgupta, *et al.,* A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **50**, 809-824 (2003).

12. A. S. Alsohim, *et al.,* The biosurfactant viscosin produced by *Pseudomonas fluorescens* SBW25 aids spreading motility and plant growth promotion. *Environ. Microbiol.* **16**, 2267-2281 (2014).

13. W. Li, C.-D. Lu, Regulation of carbon and nitrogen utilization by CbrAB and NtrBC two-component systems in *Pseudomonas aeruginosa*. *J. Bacteriol.* **189**, 5413-5420 (2007).

14. D. J. Studholme, R. Dixon, Domain architectures of σ^{54}-dependent transcriptional activators. *J. Bacteriol.* **185**, 1757-1767 (2003).

15. R. Wassem, E. M. de Souza, M. G. Yates, F. D. Pedrosam, M. Buck, Two roles for integration host factor at an enhancer-dependent *nifA* promoter, *Mol. Microbiol.* 35, 756-764 (2000).

16. E. Amiel, R. R. Lovewell, G. A. O'Toole, D. A. Hogan, B. Berwin, *Pseudomonas aeruginosa* evasion of phagocytosis is mediated by loss of swimming motility and is independent of flagellum expression. *Infect. Immun.* **78**, 2937-2945 (2010).

17. R. León, G. Espín, *flhDC*, but not *fleQ*, regulates flagella biogenesis in *Azotobacter vinelandii*, and is under AlgU and CydR negative control. *Microbiology* **154**, 1719-1728 (2008).

18. M. E. Wand, e*t al., Helicobacter pylori* FlhB function: The FlhB c-terminal homologue HP1575 acts as a “spare part” to permit flagellar export when the HP0770 FlhBCC domain is deleted. *J. Bacteriol.* **188**, 7531-7541 (2006).

19. H. E. Hoekstra, J. A. Coyne, The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995-1016 (2007).

20. G. A. Wray, The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev. Genet.* **8**, 206-216 (2007).

21. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25-36 (2008).

22. M. W. Silby, *et al*., Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens. Genome Biol.* **10,** R51 (2009).

23. F. Van Immerseel, *et al*., *Salmonella Gallinarum* field isolates from laying hens are related to the vaccine strain SG9R. *Vaccine* **31**, 4940-4945 (2013).

24. K. Rutherford, *et al.*, Artemis: sequence visualization and annotation. *Bioinformatics* **16,** 944-945 (2000).

25. D. B. Roche, M. T. Buenavista, S. J. Tetchner, L. J. McGuffin, The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. *Nucleic Acids Res.* **39**, W171-W176 (2011).

26. L. J. McGuffin, M. T. Buenavista, D. B. Roche, The ModFOLD4 server for the quality assessment of 3D protein models. *Nucleic Acids Res.* **41**, W368-W372 (2013).

27. L. J. McGuffin, Intrinsic disorder prediction from the analysis of multiple protein fold recognition models. *Bioinformatics* **24**, 1798-1804 (2008).

28. D. B. Roche, M. T. Buenavista, L. J. McGuffin, The FunFOLD2 server for the prediction of protein–ligand interactions. *Nucleic Acids Res.* **41**, W303-W307 (2013).

29. Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302-2309 (2005).

Acknowledgments:

TBT, LJJ, RWJ and MAB conceived and designed the study. TBT, GM and AA performed experiments. MWS and AHD performed independent lines of enquiry on Pf0-2x. DS conducted bioinformatics analysis of genome resequencing data, identified mutated genes and handled sequencing data. LM conducted the protein structure prediction and analysis. This work was supported by a Leverhulme grant to LJJ, MAB and RWJ, BBSRC grant BB/J015350/1 to RWJ, start-up funding from the University of York to MAB, Qassim University to AA, and Agriculture and Food Research Initiative Competitive Grant 2010-65110-20392 from the USDA's National Institute of Food and Agriculture, Microbial Functional Genomics Program to MWS. TBT, GM, LJJ, RWJ, MWS, DJS and MAB wrote the paper. We thank Graham Bell, Mark Pagel, Angus Buckling and James Moir for useful discussions; Peter Ashton for processing of microarray data; and Konrad Paszkiewicz and Exeter Sequencing Service facility and support from the following: Wellcome Trust Institutional Strategic Support Fund (WT097835MF), Wellcome Trust Multi User Equipment Award (WT101650MA) and BBSRC LOLA award (BB/K003240/1). Sequence data for genomic resequencing of AR2S and AR2F have been submitted to the SRA under accession numbers SRR1510202 and SRR1510203, respectively. The eArray design ID for the microarray is 045642. Microarray data have been submitted to the ArrayExpress database under accession number E-MTAB-2788 (www.ebi.ac.uk/arrayexpress).

Figures 1 – 3:

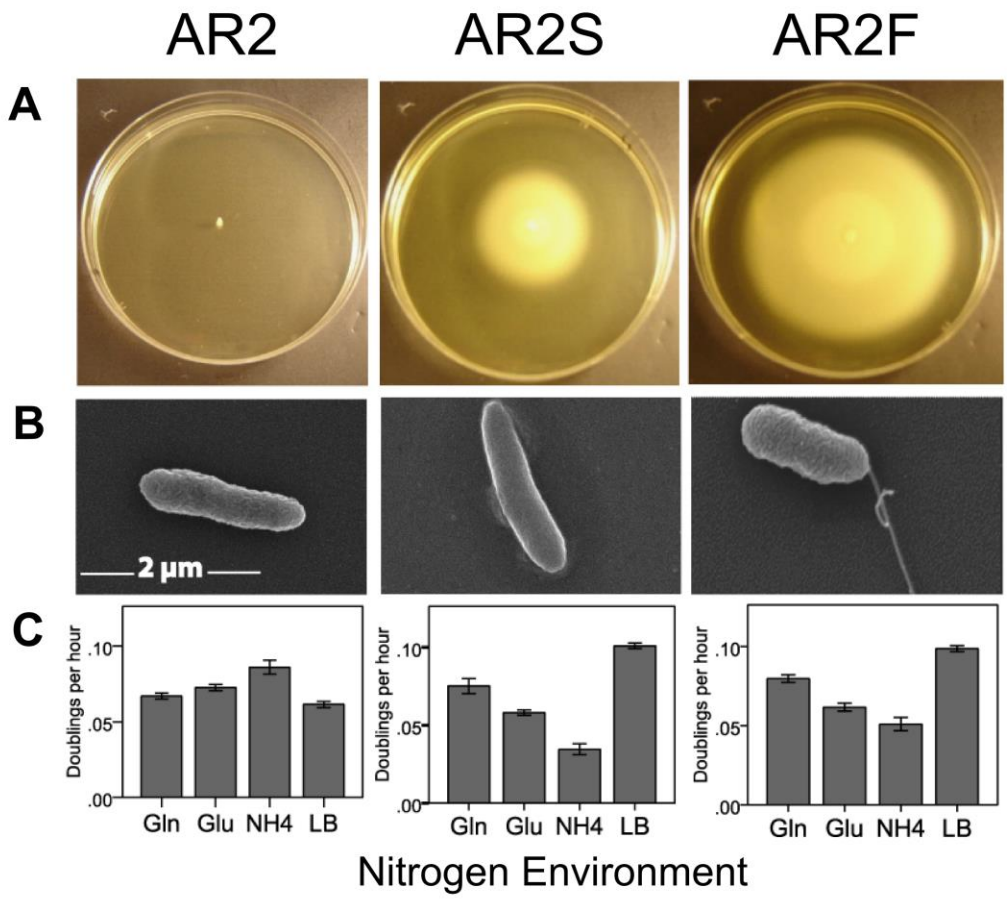


Fig. 1. **Phenotypic assays of motility variants.**

A Surface spreading motility assays of ancestral (AR2) and evolved slow spreading (AR2S) and fast spreading (AR2F) mutants, after 27 hours. **B** Electron microscopy confirms the presence of a flagellum in AR2F, but fails to confirm presence in AR2S. **C** Mean (N=4) cell doublings per hour (+/- 1 SEM). Strains were grown within differing nitrogen environments: 10 mM glutamine (Gln), glutamate (Glu) and ammonium (NH4) as the sole nitrogen source, or in high nutrient lysogeny broth (LB): AR2, $F_{3, 12} = 13.460$, $P < 0.001$; AR2S, $F_{3, 12} = 72.674$, $P < 0.001$; AR2F, $F_{3, 12} = 52.538$, $P < 0.001$. There were also differences between doubling rate of strains within each nitrogen medium (Glutamate (Glu), $F_{2, 9} = 12.654$, $P = 0.002$; Ammonium (NH4), $F_{2, 9} = 40.529$, $P < 0.001$), with the exception of Glutamine (Gln) ($F_{2, 9} = 3.703$, $P = 0.067$).

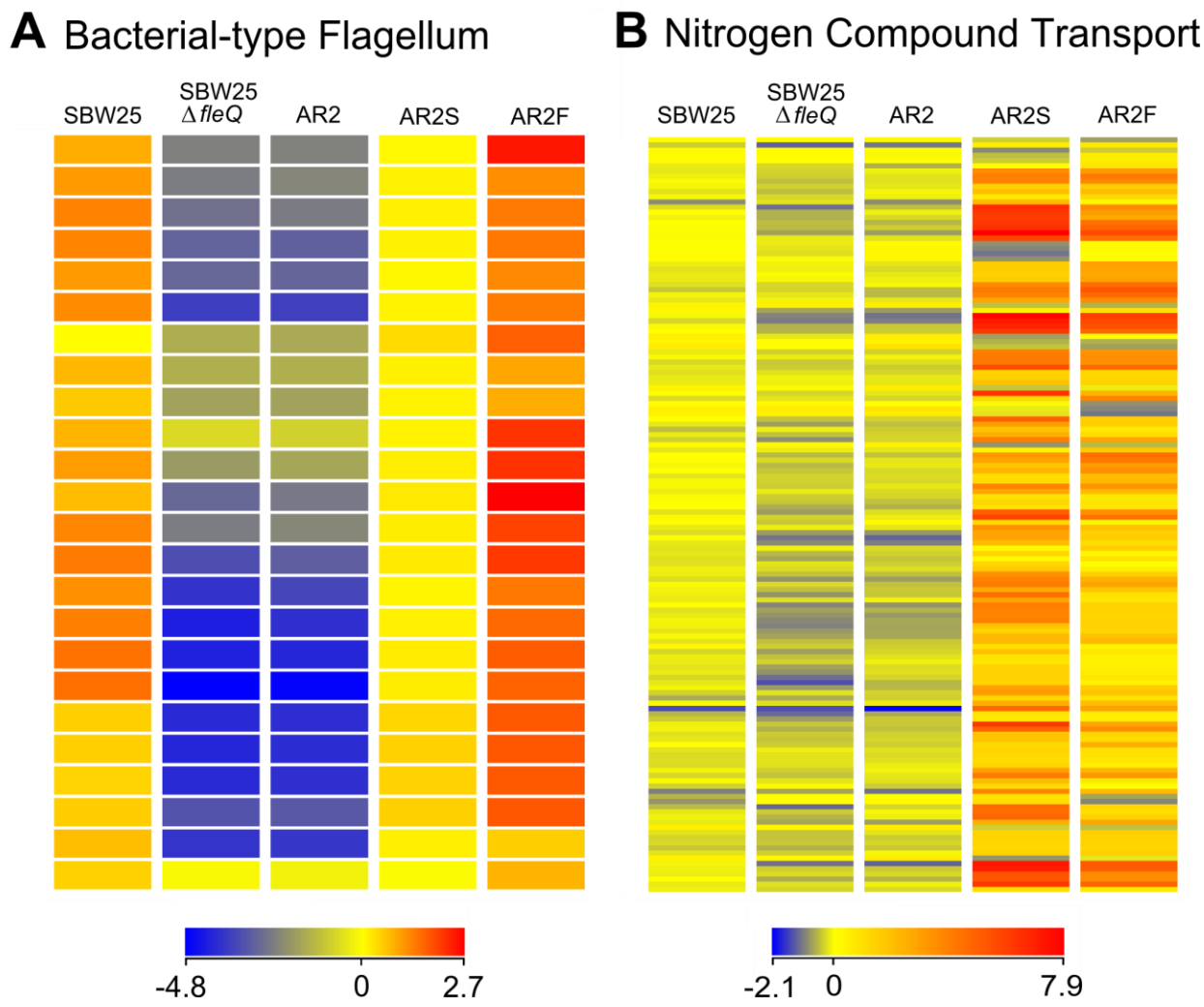


Fig. 2. Heat map of microarray expression profiles for all evolved and ancestral motility variants.

Heat maps show where there is significant ($P \leq 0.05$) fold-change of ≥ 2 in genes selected based on GO-terms for **A** Bacterial-type flagellum (24 genes) and **B** Nitrogen compound transport (146 genes) for all strains. The gradation of colors reflects normalized raw signal values across the entire array. Genes are ordered according to chromosomal position to enable clearer visualization of coregulated gene clusters. Full transcriptome analysis is reported in Supplementary File “Microarray dataset.xlsx”.

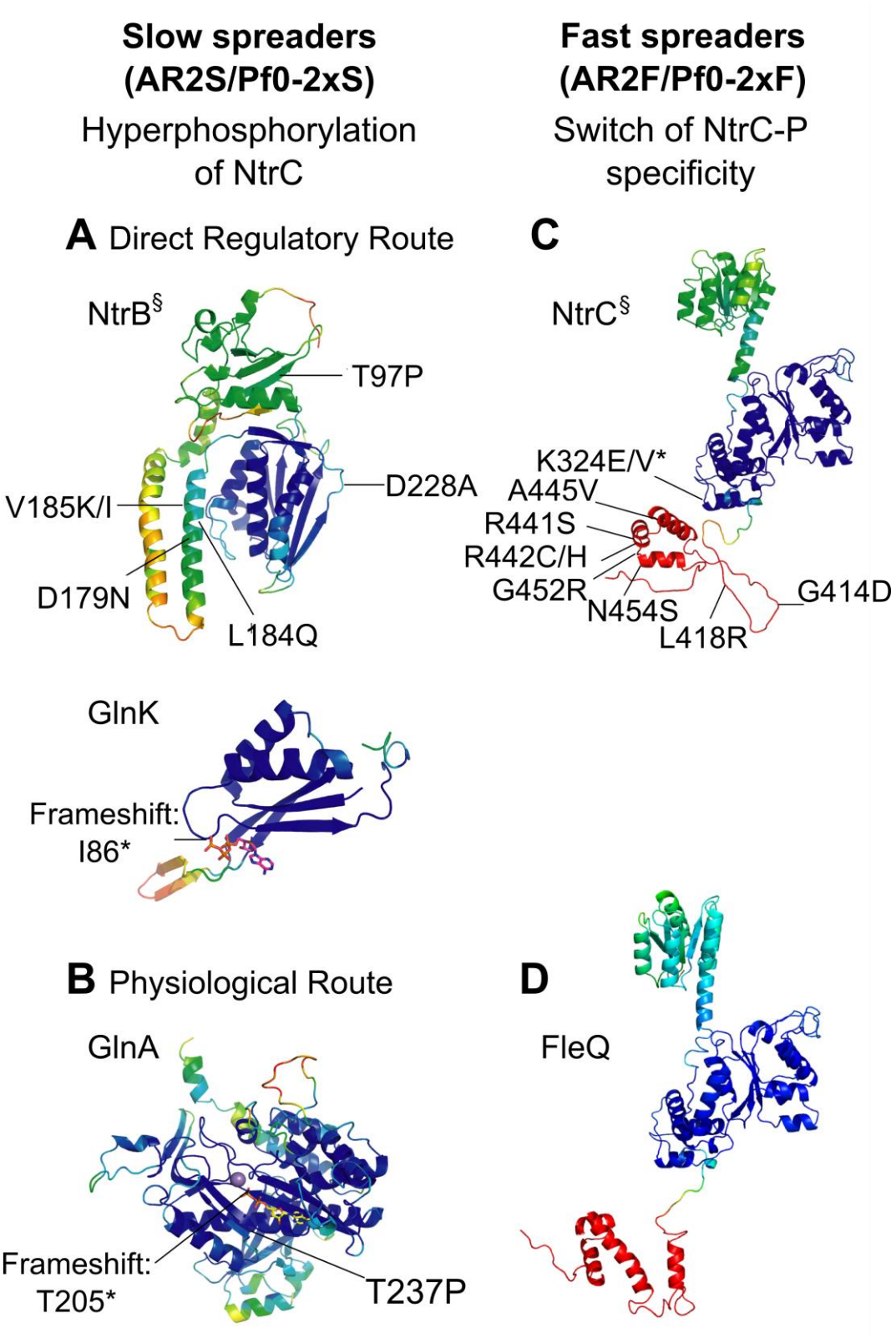

Slow spreaders
(AR2S/Pf0-2xS)
Hyperphosphorylation
of NtrC
Fast spreaders
(AR2F/Pf0-2xF)
Switch of NtrC-P
specificity
A Direct Regulatory Route
NtrB§
T97P
D228A
V185K/I
D179N
L184Q
GlnK
Frameshift:
I86*
B Physiological Route
GlnA
Frameshift:
T205*
T237P
C
NtrC§
K324E/V*
A445V
R441S
R442C/H
G452R
N454S
G414D
L418R
D
FleQ

Fig. 3. Full chain multi-template 3D models of protein structures of slow and fast spreading motility variants.

Slow spreading variants can either follow the direct regulatory route through mutation of NtrB or GlnK (**A**), or the physiological route through mutation of GlnA causing over-activation of NtrB (**B**); both routes are predicted to lead to hyperphosphorylation of NtrC. Fast spreading variants all show mutational changes to NtrC (**C**), which has a similar global structure to FleQ (**D**). The colour scheme represents the variation in models, which correlates with local (per-residue) quality and disorder. Regions coloured in blue and green represent low local variability in structure, while those in red show high local variability (see Table 1 and Table S2 for mutation details). § = All mutations mapped onto SBW25 wildtype protein structures for illustrative purposes; * = Truncated domain.

	Slow Spreaders (AR2/Pf0-2xS) Hyperphosphorylation of NtrC	**Fast Spreaders (AR2/Pf0-2xF)** Switch of NtrC-P specificity
AR2	*ntrB* T97P*	*ntrC* R442C*
	ntrB V185K	*ntrC* K342E
	ntrB D179N	*ntrC* G452R
	ntrB L184Q / V185I	*ntrC* K342V / Frameshift: V342
Pf0-2x	*ntrB* D228A§	*ntrC* N454S
		ntrC R441S
		ntrC N454S
		ntrC P424L
		ntrC N454S
	ntrB D228A§	*ntrC* L418R
		ntrC G414D
		ntrC N454S
		ntrC F426V
	glnK Frameshift: I86*	*ntrC* R442H*
	glnA T237P*	*ntrC* A445V*
	glnA Frameshift: T205*	*ntrC* R442C*

* = Identified by genome resequencing

§ = Independent *ntrB* mutant strains, parent to multiple *ntrC* mutant strains

Table 1. Mutational trajectory towards slow and fast spreading phenotypes.

Mutations confirmed in slow spreading motility variants are predicted to result in hyperphosphorylation of NtrC; mutations in fast spreading variants lead to predicted switched specificity of NtrC-P towards FleQ targets. Slow and fast spreading variants share the same ancestry.

Supplementary Materials:

Materials and Methods

Fig. S1.

Table S1.

Table S2.

Supplementary Materials:

Materials and Methods:

Microbiological methods

P. fluorescens strains used in this study: SBW25, SBW25Δ*fleQ*, AR2 (SBW25Δ*fleQ* IS-ΩKm-hah: PFLU2552), and Pf0-2x (Pf0-1Δ*fleQ*). Motility assays are as described in (*12*). All starting populations were from a single AR2 colony grown on LB agar (1.5%), and stab inoculated into the center of a SMM plate using a sterile wire. The initial observed motility mutants (AR2S and AR2F) were cultured immediately, cryopreserved and used for subsequent genome resequencing and microarray analysis. Independently evolved motility mutants were isolated from AR2 and Pf0-2x and cryopreserved. Mutation rates of all lineages were checked by plating an overnight culture onto LB agar supplemented with 100 µg ml^{-1} rifampicin. All strains were found to have a similar

mutation frequency: SBW25 = 6.98 x 10^{-9} cfu ml^{-1}; ΔfleQSBW25 = 3.72 x 10^{-8} cfu ml^{-1}; AR2 = 2.50 x 10^{-8} cfu ml^{-1}; AR2S = 2.3 x 10^{-8} cfu ml^{-1}; AR2F = 1.4 x 10^{-8} cfu ml^{-1}.

M9 nitrogen modified medium (lacking ammonium) was supplemented with 10 mM glutamate, glutamine or ammonium solutions. Strains were grown for 16 hours in LB at 27ºC and diluted to OD 0.001 in M9 nitrogen modified media, and in LB. Optical density 595 nm was measured every 20 minutes, under continuous shaking and at an incubation of 27ºC, for 24 hours (Tecan, GENios).

Molecular methods

The mutations present in the original evolved mutants AR2S and AR2F, plus 3 Pf0-2xS and 3 descended Pf0-2xF strains were identified by genome resequencing.

Genomic DNA was isolated using Puregene DNA extraction (Qiagen) or Wizard® Genomic DNA Purification (Promega) kits. We resequenced the complete genomes of mutants AR2S and AR2F (original) using the Illumina HiSeq and identified single-nucleotide variants (SNV) by alignment to the SBW25 reference as described previously (*22, 23*). Further mutants were analyzed by targeted PCR amplification and sequencing of *ntrBC* to determine whether the same two-step process had occurred; primers were designed to flank the region where SNVs had been identified by whole-genome resequencing (NtrC, F: CTTCATCCCCAACTCCTTGA, R: AAGCTGCTGAAAAGCGAGAC; NtrB, F: CTTGCGCCTTGAGTACATGA, R: ATGCGGTCTACCAGGTTACG).

AR2S and AR2F Isolate Genome Resequencing Details

Whole-genome resequencing was performed using the Illumina GA2x. We generated 12,065,035 pairs of 36-bp reads for AR2S (SRA accession SRR1510202) and 24,075,130 pairs of 36-bp reads for AR2F (SRR1510203). To identify mutations, we aligned the genomic sequence reads against the *Pseudomonas fluorescens* SBW25 genome sequence (NCBI RefSeq accession NC_012660) using BWA-mem version 0.7.5a-r405 (http://www.ncbi.nlm.nih.gov/pubmed/19451168). We discarded reads that did not uniquely map to a single location on the SBW25 genome in order to avoid artifacts due to misalignment of sequence reads arising from repeat sequences. The resulting average depths of coverage over the SBW25 genome were 124x and 144x for AR2S and AR2F, respectively. For 99.86% of the SBW25 reference genome sequence (6,713,197 out of 6,722,539 bp) we were able to unambiguously determine the nucleotide sequence in both mutant genomes (on the basis that at these genomic positions there was at least 95% consensus among all aligned reads at those positions for each of the two resequenced genomes and depth of at least 5x). Over the 6,713,197 bp of the genome for which we could unambiguously determine the DNA sequence for both mutant genomes, we found only three SNVs with respect to the reference genome sequence. Two of these three variants were present in both resequenced genomes: at position 376,439 both resequenced strains had a G whereas in the SBW25 reference genome the base is T. This corresponds to a non-silent change from codon acc to codon Ccc in gene PFLU0344 (*glnL/ntrB*). The second variant was at position 1,786,536 where SBW25 has A but the two resequenced mutants have G; this variant falls in an intergenic region. The third SNV is private to AR2F and falls within the gene PFLU0343

(*glnG/ntrC*). In SBW25 and AR2S the base is G but in AR2F it is A; this results in a non-silent change from codon cgt to codon Tgt. Aside from these three variant sites, the remainder of the 6,713,197 bp of unambiguously resolved genome sequence contained no variation from the SBW25 reference genome.

Pf0-2xS and Pf0-2xF Isolate Genome Resequencing Details

Library preparation was performed using Illumina Nextera XT kit with Nextera XT Indexes and sequenced as 250PE reads from MiSeq.

Whole-genome resequencing of Pf0-2x strains was performed using the Illumina MiSeq. We generated 1,145,122 forward and 1,141,017 reverse reads of 251-bp for Pf0-2xS_1, 1,279,360 forward and 1,272,990 reverse reads of 251-bp for Pf0-2xF_1.1, 1,962,339 forward and 1,941,676 reverse reads of 251-bp for Pf0-2xS_2, 907,328 forward and 906,083 reverse reads of 251-bp for Pf0-2xF_2.1, 935,862 forward and 932,415 reverse reads of 251-bp for Pf0-2xS_3, and 1,194,665 forward and 1,190,975 reverse reads of 251-bp for Pf0-2xF_3.1. These genomic sequence reads were mapped against the *Pseudomonas fluorescens* Pf0-1 genome sequence (NCBI RefSeq accession NC_007492.2) using CLC Genomics Workbench 6.0; unmapped reads were not included in further analysis. The average coverage depths for each library were: 44.6 for Pf0-2xS_1, 49.8 for Pf0-2xF_1.1, 76.1 for Pf0-2xS_2, 35.3 for Pf0-2xF_2.1, 36.4 for Pf0-2xS_3, and 46.5 for Pf0-2xF_3.1. Out of the 6,438,405 bp Pf0-1 genome, 97.06-99.35% of nucleotides (depending on sample) had a minimum 95% consensus among all mapped reads. Using probabilistic variant detection to investigate unambiguous nucleotide positions in each sequenced strain, we identified two SNVs and a 3bp

deletion common to all mutants (Table S2). These three common changes were also present in the parental strain (Pf0-2x) and were thus not considered further. A 15bp deletion was identified in both the Pf0-2xS1 and F1.1 strains with a unique SNV occurring in the latter. An additional SNV was common to Pf0-2xS2 and Pf0-2xF2.1, yet another to both Pf0-2xS3 and Pf0-2xF3.1, and a single unique SNV was discovered in Pf0-2xF2.1 and another in Pf0-2xF3.1 with respect to the reference genome sequence.

The presence of *glnA*, *glnK* and *ntrC* mutations in resequenced strains was confirmed by PCR of candidate regions from parental and evolved strains followed by Sanger sequencing of the amplicons. Mutations in *ntrB* were detected by targeted PCR amplification and sequencing, and *ntrC* mutations in derivatives of *ntrB* mutants were detected the same way. In some cases, multiple primers were used to cover the full length of the gene. This enabled us to be certain about the mutations, and rule out any others. Primers used for Pf0-1 were: ntrB, F: ATTGCGCCTCGAGTACATGA and TCGACACGGTTTCACTACGG, R: ATGCGGTCGACCAGATTGCG and TCGGAGCCTTGGTTTGGTTT; ntrC, F: AAGCTGCTGAAAAGCGAGAC and GATTAAGGGTCACGGTGCCT, R: CTTCATGCCGAGTTCCTTGA and CACTGGAACAAGGAGCCACA; glnA, F: GGAGGCCTTTCTTTGTCACG and ACGCTTGTAGGAGTTGGTGG and CGGGAAGTAGCCACCTTTCA, R: CCACCAACTCCTACAAGCGT and CAAGTCCGACATCTCCGGTT and CTACCCGCCCTAATTCACCC; glnK, F: GCCGGGCATTACGATAGACA, R: TGCGCTTAGACTTGAGTCGG.

Microarray design

For microarray analysis, RNA was extracted from SBW25, SBW25Δ*fleQ*, AR2, AR2S and AR2F using an RNeasy kit (Qiagen) and assayed for quality [2100 Bioanalyzer (Agilent), and Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific)]. RNA was labelled, hybridised to custom Agilent arrays and scanned according to the manufacturer's instructions (Technology Facility, Dept. of Biology, University of York, UK). Differentially expressed genes were identified by ANOVA using the Benjamini-Hochberg FDR correction. No commercially available microarrays were available for *P. fluorescens*, so a custom design was created using the Agilent eArray system (http://earray.chem.agilent.com/earray). The *P. fluorescens* SBW25 genome was loaded into Artemis (*24*) (http://www.sanger.ac.uk/resources/software/artemis), and all open reading frames marked as CDSs and these putative CDS written to a FASTA file which was uploaded to eArray. A probeset was created containing five unique probes per CDS, plus the appropriate Agilent control sequences. The probes were then laid out onto a standard Agilent 8x60K format slide. Microarrays were validated using qRT-PCR (MyiQTM, Bio-Rad).

RNA labelling and hybridisation

RNA isolated from each sample was labelled with Cy-3 using the Agilent Low Input Quick Amp WT Labelling Kit (one color) according to the manufacturer's instructions. Briefly, the RNA is reverse transcribed using a primer mix containing oligo-dT and random nucleotide primers containing T7 promoters, then the cRNA amplified using the T7 polymerase in the presence of Cy-3 labelled dCTP. The labelled cDNA was then

quality controlled and only samples with a specific activity of > 6 were used hybridised to the arrays.

Samples were hybridised to the arrays, incubated and washed in accordance with the manufacturer's instructions, and the slides scanned using an Agilent Microarray Scanner and the fluorescence quantified with the Feature Extraction software. The resulting files were then loaded into GeneSpring for further analysis.

Initial Data Analysis

GeneSpring software version 12.6 was used for all data analysis. Initial analysis was performed to identify those genes that were differentially expressed between any conditions in the experiment, followed by Gene Set Expression Analysis (GSEA) on the basis of Gene Ontology terms. Briefly, the sample replicates were grouped, and then ANOVA performed to identify differentially expressed genes, using the Benjamini-Hochberg FDR p-value correction (with a FDR of 5%), and a Tukey HSD post-hoc test used for each gene to identify which samples were significantly different from each other for that gene. Those GO categories that were over-represented in the set of differentially expressed genes were then identified.

3D protein models

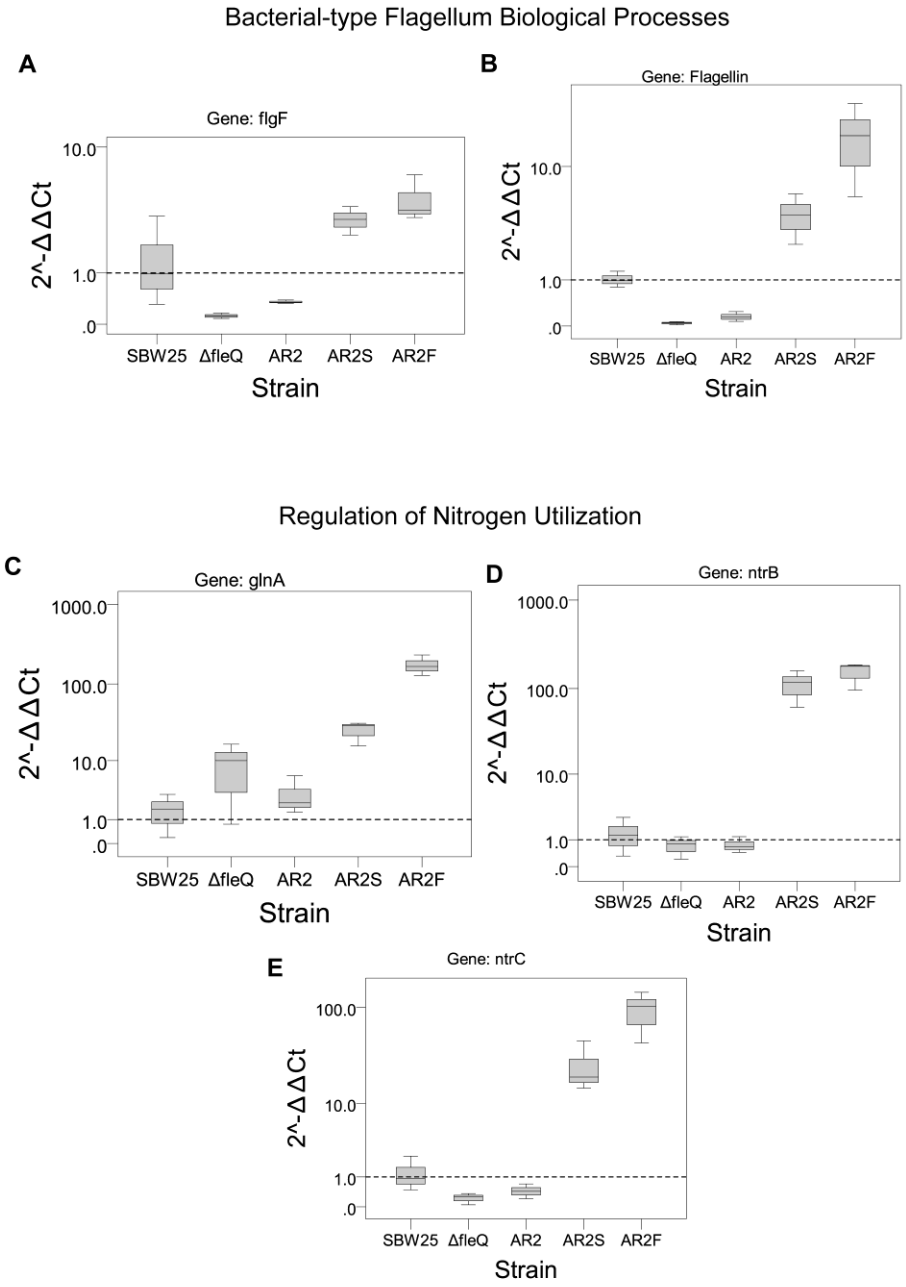
The IntFOLD server (*25*) version 2 was used to build multi-template 3D models from the wild type and mutant sequences of AdnA, NtrB, NtrC, FleQ, GlnA and GlnK. The 3D models were quality checked using the ModFOLD4 protocol (*26*). Intrinsic protein

disorder was predicted using the DISOclust method (*27*). Binding sites were predicted using the FunFOLD2 method (*28*). Models were structurally aligned and scored using TM-align (*29*).

Further Discussion of Microarray Results

Full transcriptome analysis is reported separately in Supplementary File “Microarray dataset.xlsx”. When genes were filtered for significant fold changes ($P \leq 0.05$) of ≥ 2, and sorted by descending fold changes, the genes that showed the greatest changes in expression are those highlighted in this study. Specifically, when comparing AR2 and AR2S, the large majority of genes with greatest up regulation are related to nitrogen transport, whereas when comparing AR2S to AR2F, the large majority of genes are related to flagellar motility and chemotaxis.

Our microarray data show that the average level of flagella gene expression in AR2S is between 1.5-2-fold lower than in SBW25, but considerably higher than AR2 (between 3 and 75 fold). There are at least two possible models to explain why we were unable to observe flagella in AR2S: (i) Phosphorylated NtrC (NtrC-P) only weakly interacts with flagella gene promoters, but the over-abundance of NtrC-P in AR2S saturates NtrC-dependent promoters and the excess NtrC-P drives expression of flagella genes. The level of transcription is suboptimal, thus the flagella may be unstable and support limited or transient motility; (ii) Promiscuous activity of over-abundant NtrC-P (e.g. interacting with RpoN from solution) leads to expression from flagella promoters, but this is not equivalent in all cells and only a proportion of the population express sufficient levels to produce a functional flagellum at any one time.



372

Fig. S1: **Validation of gene expression changes detected by microarray using qRT-PCR.**

A – E Box plots show gene expression, relative to wild-type SBW25, in genes related to nitrogen utilization and flagellar function. Note, data presented on log scales.

A. SBW25Δ*fleQ*

	GENE SYMBOL	GENE CHIP ARRAY		REAL TIME PCR	
		Fold Change	p-value	Fold Change	p-value
Bacterial-type Flagellum	*flgF*	-79.06	1.52E-07	-8.4	0.196
	Pflu4448 (Flagellin)	-158.33	2.09E-07	-22.59	0.002
Regulation of Nitrogen Utilization	*glnA*	1.01	2.24E-06	9.14	0.187
	ntrB	-1.54	1.37E-07	-4.5	0.41
	ntrC	-1.74	1.58E-07	-1.39	0.136

B. AR2

	GENE SYMBOL	GENE CHIP ARRAY		REAL TIME PCR	
		Fold Change	p-value	Fold Change	p-value
Bacterial-type Flagellum	*flgF*	-76.48	1.52E-07	-2.84	0.265
	Pflu4448 (Flagellin)	-128.22	2.09E-07	-6.61	0.004
Regulation of Nitrogen Utilization	*glnA*	1.03	2.24E-06	3.27	0.393
	ntrB	-1	1.37E-07	-2.24	0.424
	ntrC	-1.41	1.58E-07	-1.31	0.231

C. AR2S

	GENE SYMBOL	GENE CHIP ARRAY		REAL TIME PCR	
		Fold Change	p-value	Fold Change	p-value
Bacterial-type Flagellum	*flgF*	-2.53	1.52E-07	3.13	0.192
	Pflu4448 (Flagellin)	-1.71	2.09E-07	4.34	0.042
Regulation of Nitrogen Utilization	*glnA*	6.08	2.24E-06	25.49	0.008
	ntrB	28.18	1.37E-07	26.45	0.017
	ntrC	27.14	1.58E-07	49.14	0.059

D. AR2F

	GENE SYMBOL	GENE CHIP ARRAY		REAL TIME PCR	
		Fold Change	p-value	Fold Change	p-value
Bacterial-type Flagellum	*flgF*	1.1	1.52E-07	4.49	0.1
	Pflu4448 (Flagellin)	3.6	2.09E-07	16.57	0.065
Regulation of Nitrogen Utilization	*glnA*	11.34	2.24E-06	175.63	0.005
	ntrB	58.27	1.37E-07	96.09	0.006
	ntrC	63.72	1.58E-07	49.73	0.03

Table S1: Mean fold change (relative to SBW25) from qRT-PCR and microarrays with p-values (**A – D**).

Colours correspond to genes with significant (≥ 2 fold change) up regulated (red) and down regulated (blue) changes.

Strain	Nucleotide change	AA change	Gene	Domain/Function
AR2S_1	T376439G*	T97P	*ntrB*	PAS domain
AR2F_1.1	G374322A*	R442C	*ntrC*	HTH domain
AR2S_2	G376175A/ T376174A	V185K	*ntrB*	HK domain
AR2F_2.1	T374622C	K342E	*ntrC*	HTH domain
AR2S_3	G376193A	D179N	*ntrB*	HK domain
AR2F_3.1	C374291G	G452R	*ntrC*	Miscellaneous function
AR2S_4	G376176A T376174A	L184Q V185I	*ntrB*	HK domain
AR2F_4.1	T374622C/ *I* A374625	K342V *FS* from V342	*ntrC*	HTH domain
Pf0-2xS_1	Δ6431771-3* Δ6171472-87*	ΔG85 *FS* from I86	*gidB* *glnK*	Methyltransferase domain PII regulator
Pf0-2xF_1.1	C382879T*	R442H	*ntrC*	HTH domain
Pf0-2xS_2	Δ6431771-3* T388861G* Δ902225*	ΔG85 T237P *FS* from S544	*gidB* *glnA* *carB*	Methyltransferase domain Catalytic domain Miscellaneous function
Pf0-2xF_2.1	G382870A*	A445V	*ntrC*	HTH domain
Pf0-2xS_3	Δ6431771-3* Δ388958*	ΔG85 *FS* from T205	*gidB* *glnA*	Methyltransferase domain Catalytic domain
Pf0-2xF_3.1	G382880A*	R442C	*ntrC*	HTH domain
Pf0-2xS_4	A384603C	D228A	*ntrB*	Miscellaneous function
Pf0-2xF_4.1	A382843G	N454S	*ntrC*	HTH domain
Pf0-2xF_4.2	C382883A	R441S	*ntrC*	HTH domain
Pf0-2xF_4.3	A382843G	N454S	*ntrC*	HTH domain
Pf0-2xF_4.4	C382933T	P424L	*ntrC*	HTH domain
Pf0-2xF_4.5	A382843G	N454S	*ntrC*	HTH domain
Pf0-2xS_5	A384603C	D228A	*ntrB*	Miscellaneous function
Pf0-2xF_5.1	T382951G	L418R	*ntrC*	Miscellaneous function
Pf0-2xF_5.2	G382963A	G414D	*ntrC*	Miscellaneous function
Pf0-2xF_5.3	A382843G	N454S	*ntrC*	HTH domain
Pf0-2xF_5.4	T382928G	F426V	*ntrC*	HTH domain

* = Identified by genome resequencing.

Table S2: **Full details of mutations identified in strains with parental AR2 and Pf0-2x origin.**

Numbers before decimal points represent independent lineages, and numbers after represent derived strains. Note fast spreading mutants contain all nucleotide changes within a lineage.